Biochemical Genetics, Vol. 45, Nos. 1/2, February 2007 (© 2007)
DOI: 10.1007/s10528-006-9048-8

Note

Evolution of the *Spiroplasma* P58 Multigene Family

Jana Comer,[1] Jacqueline Fletcher,[2] Robert E. Davis,[3] and Ulrich Melcher[1,4]

Received 9 January 2006—Final 18 May 2006
Published online: 13 January 2007

INTRODUCTION

Spiroplasma citri and *S. kunkelii* are prokaryotic phytopathogens transmitted primarily by leafhopper vectors. *S. citri* BR3-3X was originally isolated from horseradish plants with brittle root disease (Fletcher *et al*., 1981). Extensive chromosomal rearrangement was found in one derivative of *S. citri* BR3-3X after lengthy maintenance (Ye *et al*., 1996). This derivative, BR3-G, was maintained by plant-to-plant transmission and has lost insect transmissibility. Among the changes in BR3-G as compared to BR3-3X are a genomic inversion and two deletions, one at each inversion border. A 9.6 kb segment of BR3-3X corresponding to one of the areas deleted from BR3-G was sequenced (Fig. 1). ORFs in the deleted region are predicted to encode proteins of 58, 12, 54, and 123 kDa. The protein of 58 kDa has limited amino acid sequence similarity with known mycoplasma adhesins (Ye *et al*., 1997).

In addition to the characterized *S. citri* P58 gene (ScA), at least two other P58-like sequences (ScB, ScC) are present in *S. citri* BR3-3X (Ye *et al*., 1997). Southern blot analysis using a probe corresponding to the 5′ region of P58 revealed three reactive *Eco*RI fragments in the insect transmissible lines *S. citri* BR3-3X and BR3-T, but only two in the insect nontransmissible line BR3-G (unpublished data). The observation that the characterized *S. citri* P58 gene is not present in the insect nontransmissible line *S. citri* BR3-G suggests that the P58 multigene family may play a role in insect interaction. However, because a P58 protein was detected in BR3-G despite the deletion of the characterized P58 gene in this line, we know that in BR3-G at least one other P58-like gene is transcribed and its

[1]Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74078.
[2]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma 74078.
[3]Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland 20705.
[4]To whom correspondence should be addressed; e-mail: u-melcher-4@alumni.uchicago.edu.

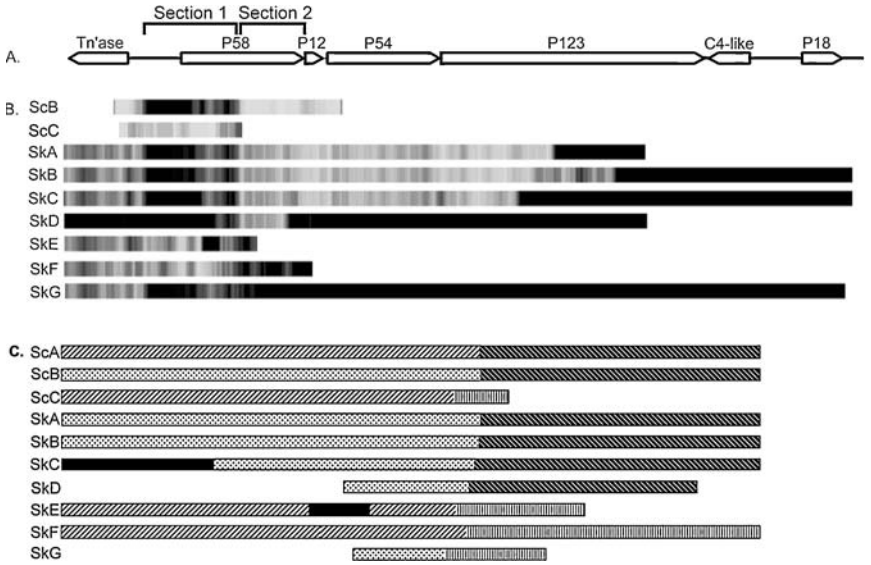

Fig. 1 *S. citri* and *S. kunkelii* P58 and flanking region and recombination junctions. (A) *S. citri* BR3-3X genome section containing P58 and adjacent ORFs. (B) P58-like genes and surrounding sections showing relative degree of similarity to *S. citri* sequence. A multiple sequence alignment of the *S. citri* and *S. kunkelii* CR2-3X P58 sequences was imported into MAViT (Multiple Alignment Visualization Tool, www.biochic.com). MAViT uses a sliding window of 100 nucleotides to calculate a percent identity between each pair of sequences of a multiple sequence alignment. Identity equal to or less than 50% is shown in black. From 50 to 100%, the shading ranges from dark gray to white. (C) Regions shaded similarly had greater than 70% identity at the nucleotide level. The two types of 5′ P58 regions are A (dark diagonal stripes) and B (dots). Solid bars indicate a deletion relative to ScA. The two types of 3′ regions are X (white diagonal stripes) and Y (vertical stripes).

mRNA is translated (Ye *et al*., 1997). Thus, the role of P58 in insect interactions remains unclear.

Because spiroplasma genomes have evolved to be reduced in size, retention of a multigene family may indicate a significant function for these genes. Multiple similar sequences may also be the result of multiple inserted viral sequences. In this study, we analyzed the *S. citri* and *S. kunkelii* P58 multigene family and flanking sequences (Fig. 1). As expected, regions of high identity were found among the P58 family members. Regions of low-percentage identity, however, between some of the P58 family members were discovered in the 5′ portion of the P58 sequences. These regions of low identity and the surrounding regions of high identity were analyzed. Alternate sequence types for the 5′ (designated A or B) and 3′ (X or Y) portions of the P58-like sequences are proposed. A region in which one or more recombination events occurred, including that of an adhesin and a bacteriophage gene, may have had a higher rate of recombination. To our

knowledge, this is the first report of a phage-adhesin fusion gene in spiroplasmas. Recombination events, such as those described here, could provide a mechanism for variation in spiroplasma evolution and in niche adaptation.

METHODS

Using the *S. citri* predicted amino acid P58 sequence (accession no. AAB00186.1) as query, *S. kunkelii* sequences selected for study from *S. kunkelii* CR2-3X genome data (Zhao *et al*., 2003) had expected values less than 1×10^{-50} from a tBLASTn search performed on 7-15-04 (www.genome.ou.edu/spiro.html). For ease of analysis, large retrieved *S. kunkelii* contigs (>10 kb) were trimmed to no closer than 4 kb to the P58 region. Contigs are designated herein by letters of the alphabet (SkA-SkG).

S. citri BR3-G was grown to exponential phase in LD8 broth (Chen and Davis, 1979) at 31°C, and DNA was isolated using the CTAB method, as described by Berg *et al*. (2001). Regions on either side of or within the P58 region were used to design primers for ScB and ScC. The forward primer 5′-CCATTTTTCTTTTTACAA-3′ was used in the primer pair for both ScB and ScC. The reverse primers 5′-TTTACCCATTAAACTTGC-3′ and 5′-ACAGCACCTTGTACTTGT-3′ were used to amplify the ScB and ScC P58 sequences, respectively.

PCR of the ScB region was performed in a 50 μL reaction volume with final concentrations as follows: 1 × reaction buffer (Promega), 200 μM dNTPs, 3.5 mM $MgCl_2$, 0.485 μg *S. citri* BR3-G DNA for ScB, 0.2 μM primers, and 2.5 U *Taq* (Promega). The reaction had an initial denaturation step of 1 min at 95°C, then 35 cycles as follows: 94°C 30 s, 42°C 1 min, 72°C 3 min. A final elongation step of 5 min was performed at 72°C. PCR of the ScC region was identical with the following exceptions: 1.12 μg *S. citri* BR3-3X DNA was used as template, elongation at 72°C was 1 min 30 s, and the final elongation step was 10 min. The PCR products were cloned using a Topo TA cloning kit (Invitrogen) and sequenced by primer walking.

Only regions sharing greater than 70% identity with at least one other P58-like sequence were included in the comparisons. The positions of junctions between regions of high (>70%) and low (<70%) identity are indicated in Fig. 1. The pairwise percentage identities on either side of these boundaries were calculated and are shown in Fig. 2. Pairwise predicted amino acid alignments were also assessed for percentage identity in N-terminal and C-terminal regions divided using the same boundaries (data not shown).

PCR primers were chosen for validation of the positions of the recombination junctions. The following primers were specific to a recombination junction (5′-3′, f: forward, r: reverse) SkA (r) GATTTCAAAAGTCATATC, SkB (r) TTTAAGTTTGGTAAAGCG, SkC (r) GAAACTGCCATCTTCTAA, SkE

		S. citri			*S. kunkelii*						
		A	B	C	A	B	C	D	E	F	G
S. citri	A		59.6	95.6	58.1	61.4	58.7	62.4	89.8	89.6	57.3
	B	99.2		56.8	82.7	82.6	83.1	79.8	56.0	58.3	82.7
	C	68.5	69.9		61.0	61.1	58.6	59.7	90.0	90.0	61.0
S. kunkelii	A	89.9	90.4	65.8		99.7	93.7	91.9	57.6	60.7	99.2
	B	90.0	90.1	65.8	99.2		93.8	92.3	57.8	61.9	99.0
	C	88.3	88.9	64.4	95.4	95.4		88.1	41.1	65.9	93.8
	D	90.1	90.9	66.0	99.6	99.5	97.5		63.7	62.3	91.5
	E	66.7	62.2	83.6	61.7	61.7	36.9	61.0		96.7	58.4
	F	60.2	60.0	83.6	58.7	59.9	58.3	59.7	95.0		61.6
	G	51.7	52.5	83.6	51.5	56.5	50.1	52.5	96.3	97.6	

Fig. 2 P58 nucleotide percent identity of 5′ and 3′ sections and pairwise nucleotide sequence alignment of *S. citri* and *S. kunkelii*. The 5′ and 3′ sections of P58 were analyzed separately. Section 1, corresponding to part of the intergenic region 5′ of P58 and the 5′ half of P58 (see Fig. 1), is in the upper right, and section 2, corresponding to the 3′ half of P58, is in the lower left. Sequences with percent identities higher than 70% are shown in gray.

(f) TAAAGATATAAATTTTC, SkF (f) AAAATAAATGAGGTAATA, SkG (f) ATTTGTATCATTGCTAGT. The primers SkABC (f) GAAGAGAAGAATGTGACC and SkEFG (r) ATCTGCTTGAGCATAATC were shared among P58 sequences. PCR reactions were carried out as described above with the following exceptions: 3.0 mM $MgCl_2$ and 0.1 μg *S. kunkelii* CR2-3X DNA. Thermocycler conditions were as previously described, except for the use of 24 cycles with annealing temperatures as follows: SkA and SkG, 42°C; SkB and SkC, 45°C; SkE and SkF, 38°C.

RESULTS AND DISCUSSION

In this paper, we have designated sequences having similarity to P58 as regions, and the contigs containing P58-like regions by letters of the alphabet (ScA-ScC, and SkA-SkG). Flanking regions are sequences immediately surrounding the P58 region. The P58 region has been divided into section 1 and section 2 (see Fig. 1).

Two *S. citri* P58-like contigs, ScB and ScC, were cloned and sequenced (acc. nos. DQ344811 and DQ344812, respectively) and seven *S. kunkelii* P58 contigs

(SkA-SkG) were detected in the *S. kunkelii* genome sequence (Fig. 1). Expected values of a tBLASTn search using the ScA P58 predicted amino acid sequence as a query ranged from 3×10^{-55} (SkE) to 1×10^{-180} (SkB).

To corroborate the sequence assembly (Zhao *et al*., 2003), PCR across the putative recombination junctions of the P58 regions yielded products of expected size for contigs SkA, SkB, SkC, SkE, SkF, and SkG. Due to the length and the number of polymorphisms in the P58 region of SkD, we were unable to design a specific primer for SkD. End sequencing confirmed computer assembly.

All P58-like gene sequences were not uniformly similar to each other, however. We examined subsections of the sequence at both the nucleic acid and amino acid levels (Fig. 1). As shown in Fig. 2, section 1 (− 357 to 673) of contigs ScA and ScC and two of the *S. kunkelii* contigs, SkE and SkF, shared 89.6–96.7% identity. Conversely, contigs ScB, SkA, SkB, SkC, and SkG shared considerable identity (79.8–99.7%) with one another, but were much less similar (41.1–65.9%) to ScA, ScC, SkE, and SkF. Section 2 (746–1484) (Fig. 1) showed a different pattern of conservation among several members of the multigene family. In this region, the P58 gene ScA is more similar to the corresponding parts of contigs ScB, SkA, SkB, SkC, and SkD (88.3–99.2%) than to those of ScC, SkE, SkF, and SkG (51.7–68.5%). ScC and SkE are truncated within section 2, and their similarity with the other contigs can be assessed only to the point of the truncation. Within section 2, SkG shared 83.6–97.6% identity with ScC, SkE, and SkF in a segment of 261 nucleotides. SkA, SkB, and SkC shared sequence similar to ScA 3′ of the P58-like region. The region of similarity extended into the P12, P54, and P123 ORFs of the SkA, SkB, and SkC contigs. The predicted amino acid sequence corresponding to the coding portion of section 1 and section 2 revealed the same pattern of conservation as that of the nucleic acid sequence.

The differences in similarities between P58 5′ and 3′ sections are consistent with a previous recombination event. For discussion purposes, we represent the P58 region of ScA as AX, having an A type 5′ end (Fig. 1) and an X type 3′ end. ScB, SkA, SkB, SkC, and SkD are represented as BX, with a B type 5′ end and an X type 3′ end. ScC, SkE, and SkF are termed AY, with a 5′ end similar to that of ScA but a different 3′ end, designated Y. SkG has a B type 5′ end and part of the Y type 3′ end, and thus is designated BY. Thus, the P58 family members characterized in this study fall into four groups: AX, BX, AY, and BY. If two ancestral combinations of A or B, with X or Y recombined, then as observed, AX, BX, AY, and BY would be expected.

GeneConv (Sawyer, 1999) was used to detect probable recombination among members of the P58 multigene family. A recombination fragment between ScA and ScB was detected ($P \leq 0.05$) in the P58 coding region from position 880 past the end of the P58 coding region to the end of the sequenced region. A single polymorphism is present in this fragment at position 1243. Although not detected as a fragment by GeneConv, in the remaining portion of section 2 only

two polymorphisms were present, at positions 791 and 832. A recombination fragment was also detected in ScC from 674 to the end of the region sequenced.

Among the various P58-like sequences, the putative recombinations within the P58 gene did not occur in precisely the same place (relative to ScA, as shown in Fig. 1). The ScB, SkA, and SkB junctions were close to one another (positions 746, 745, and 740, respectively), but the SkC and SkD junctions occurred at positions 729 and 715, respectively. ScC, SkE, SkF, and SkG were at more distant positions: 674, 678, 704, and 649, respectively.

Detection of a gene conversion event among ScA, ScB, and ScC by GeneConv within the P58 coding region provides support for the hypothesis of a recombination event between groups AX and BX. Recombination events between ScA and *S. kunkelii* contigs SkA, SkB, and SkC, suggested by other data reported here (Fig. 1) were not detected by GeneConv, possibly because the *S. kunkelii* contigs are more diverged from ScA in section 2 than from ScB and ScC. *S. citri* and *S. kunkelii* contigs ScA, ScB, SkA, SkB, and SkC may have been formed via one recombination event that occurred prior to speciation of *S. citri* and *S. kunkelii*, and selective pressure may have prevented divergence between ScA and ScB and between ScA and ScC. Another possibility is that each of the contigs represents a separate recombination event. In that case, the P58 region may represent a place in the genome that has a high rate of recombination. Although SkA and SkB may have been duplicated after recombination, the occurrence of recombination junctions at slightly different sites (Fig. 1) in the other P58-like sequences supports the explanation of separate recombination events.

PSI-Blast searches were performed using one representative of each type of N-terminal and C-terminal end as query (section 1 of ScA-P58 and SkA-P58, and section 2 of ScA-P58 and SkF-P58). PSI-Blast revealed that sections 1 of both ScA-P58 and SkA-P58 are closely related to the large subunit of bacteriophage terminases (expect: ScA-P58, acc. no. NP_958178, 1×10^{-45}, iteration 4; SkA-P58, acc. no. YP_059360, 9×10^{-50}, iteration 4). The similarity of the N-terminal portion of P58 amino acid sequences (A and B) to those of bacteriophage terminases suggests that the multiple P58-like sequences may be remnants of viral sequences integrated into the *S. citri* chromosome. One would expect to find mutations in nonfunctional viral sequences, and indeed, many P58-like sequences (SkA, SkC, SkD, SkE, SkF, and SkG) had mutations that probably rendered them nonfunctional (data not shown). The P58-like sequences could have been duplicated after insertion, or may be the result of multiple viral insertion events.

PSI-Blast revealed that section 2 of ScA-P58 is related to the C-terminus (residues 331–464) of *Mycoplasma hominis* P50/Vaa lipoprotein adhesin. Cytadherence is hypothesized to be a prerequisite for mycoplasma colonization and infection (Bove, 1993). P50/Vaa in *M. hominis* is composed of a conserved N-terminus and a variable number of C-terminal repeats, which undergo

recombination, duplication, and truncation at a high frequency (Henrich *et al*., 1998). If an adhesin and bacteriophage sequence were ancestors of a recombination event, one would expect the N-terminus of the P50/Vaa adhesin homolog to be close to the bacteriophage sequence. A PSI-Blast search of the *S. kunkelii* database using *M. hominis* P50 (acc. no. CAA04901) as a query detected a 483 amino acid ORF, the N-terminal portion of which was homologous to the *M. hominis* P50 N-terminus (residues 1–106) (expect 1×10^{-5}, iteration 5). PSI-Blast searches using the C-terminal part of the 483 amino acid ORF as a query detected similarity (3×10^{-46}, iteration 3) to a bacteriophage protein (acc. no. CAD19144). The similarity of the C-terminal portion of P58 (X) to *M. hominis* P50/Vaa is likely due to a reciprocal exchange between a P50/Vaa homolog and a bacteriophage gene.

At least one of the *S. citri* P58 genes is a surface exposed membrane protein (Ye *et al*., 1997), a context unusual for a bacteriophage terminase, and the predicted transmembrane regions of both P58 and *M. hominis* P50/Vaa are in the N-terminal half (Boesen *et al*., 2001; Ye *et al*., 1997). If P58 is viral in origin, P58 may have adopted a new function, and one would expect more divergence among P58 genes than among other respective genes in *S. citri* and *S. kunkelii*. Indeed, most *S. citri* and *S. kunkelii* genes tested (25 of 32) have a high degree of sequence similarity (>80% identity; data not shown). Tailed phages, which usually encode terminases, were reported to infect spiroplasmas and mycoplasmas and insert into the chromosome (Dickinson and Townsend, 1984; Dybvig *et al*., 1987; Gourlay *et al*., 1983a,b). SpV1 viral sequences have been implicated as a major factor in the variation of the *S. citri* genome (Melcher *et al*., 1999; Ye *et al*., 1992).

If P58 does play a role in spiroplasma–insect interactions, the role could be direct, such as adhesion to the insect cell, or indirect, such as being involved in the production of a metabolite or other component necessary for transmission. Even if P58 is essential for transmission, one P58 gene may be sufficient; if so, mutations in the other P58-like genes would not affect insect transmission. *S. citri* and *S. kunkelii* are transmitted by different insect species. Because spiroplasma–insect interactions are relatively specific, it would not be surprising if adhesins differed between *S. citri* and *S. kunkelii* (Fletcher *et al*., 2005).

Spiroplasmas must adapt to very different environments in the plant host and insect vector. Thus, expressing different genes at different times is likely necessary for survival. Because of the limited *Spiroplasma* genome size, spiroplasmas probably do not have a full complement of all the genes they could possibly need in the different environments, and thus they likely rely heavily on mechanisms such as flux of genetic information (Melcher and Fletcher, 1999). If so, chromosomal recombinations such as those described here may affect the evolution of *Spiroplasma* as much as, or even more than, that of other bacteria.

REFERENCES

Berg, M., Melcher, U., and Fletcher, J. (2001). Characterization of *Spiroplasma citri* adhesion related protein SARP1, which contains a domain of a novel family designated sarpin. *Gene* **275**:57–64.

Boesen, T., Fedosova, N., Kjeldgaard, M., Birkelund, S., and Christiansen, G. (2001). Molecular design of *Mycoplasma hominis* Vaa adhesin. *Protein Sci*. **10**:2577–2586.

Bove, J. (1993). Molecular features of mollicutes. *Clin. Infect Dis*. **17**:S10–S31.

Chen, T. A., and Davis, R. E. (1979). *Cultivation of Spiroplasmas*, Academic Press, New York.

Dickinson, M. J., and Townsend, R. (1984). Integration of a temperate phage infecting *Spiroplasma citri*. *Isr. J. Med. Sci*. **20**:785–787.

Dybvig, K., Liss, A., Alderete, J., Cole, R. M., and Cassell, G. H. (1987). Isolation of a virus from *Mycoplasma pulmonis*. *Isr. J. Med. Sci*. **23**:418–422.

Fletcher, J., Melcher, U., and Wayadande, A. (2005). The phytopathogenic spiroplasmas. In Dworkin, M., *et al*. (eds.), *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, 3rd edn., release 3.XX, Springer, New York. http://141.150.157.117:8080/prokPUB/index.htm.

Fletcher, J., Schultz, G. A., Davis, R. E., Eastman, C. E., and Goodman, R. E. (1981). Brittle root disease of horseradish: Evidence for an etiological role of *Spiroplasma citri*. *Phytopathology* **71**:1073–1080.

Gourlay, R. N., Wyld, S. G., and Garwes, D. J. (1983a). Some properties of mycoplasma virus Br 1. *Arch. Virol*. **75**:1–15.

Gourlay, R. N., Wyld, S. G., and Poulton, M. E. (1983b). Some characteristics of mycoplasma virus Hr 1, isolated from and infecting *Mycoplasma hyorhinis*. Brief report. *Arch. Virol*. **77**:81–85.

Henrich, B., Lang, K., Kitzerow, A., MacKenzie, C., and Hadding, U. (1998). Truncation as a novel form of variation of the P50 gene in *Mycoplasma hominis*. *Microbiology* **144**:2979–2985.

Melcher, U., and Fletcher, J. (1999). Genetic variation in *Spiroplasma citri*. *Eur. J. Plant. Pathol*. **105**:519–533.

Melcher, U., Sha, Y., Ye, F., and Fletcher, J. (1999). Mechanisms of spiroplasma genome variation associated with SpV1-like viral DNA inferred from sequence comparisons. *Microb. Comp. Genomics* **4**:29–46.

Sawyer, S. A. (1999). GeneConv: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. http://www.math.wustl.edu/∼sawyer.

Ye, F., Laigret, F., Whitley, J. C., Citti, C., Finch, L. R., Carle, P., Renaudin, J., and Bove, J. (1992). A physical and genetic map of the *Spiroplasma citri* genome. *Nucleic Acids Res*. **20**:1559–1565.

Ye, F., Melcher, U., and Fletcher, J. (1997). Molecular characterization of a gene encoding a membrane protein of *Spiroplasma citri*. *Gene* **189**:95–100.

Ye, F., Melcher, U., Rascoe, J. E., and Fletcher, J. (1996). Extensive chromosome aberrations in *Spiroplasma citri* strain BR3. *Biochem. Genet*. **34**:269–285.

Zhao, Y., Hammond, R. W., Jomantiene, R., Dally, E., Lee, I. M., Jia, H., Wu, H., Lin, S., Zhang, P., Kenton, S., Najar, F. Z., Hua, A., Roe, B. A., Fletcher, J., and Davis, R. E. (2003). Gene content and organization of an 85-kb DNA segment from the genome of the phytopathogenic mollicute *Spiroplasma kunkelii*. *Mol. Genet. Genomics* **269**:592–602.